Molecular Cell, *Volume 43*

Supplemental Information

Integrative Regulatory Mapping Indicates that the RNA-Binding Protein HuR Couples Pre-mRNA Processing and mRNA Stability

Neelanjan Mukherjee, David L. Corcoran, Jeffrey D. Nusbaum, David W. Reid, Stoyan Georgiev, Markus Hafner, Manuel Ascano, Jr., Thomas Tuschl, Uwe Ohler and Jack D. Keene

Supplemental Experimental Procedures

Filtering Ago and HuR PAR-CLIP clusters. Clusters were required to contain conversions at two or more locations to reduce noise and potential false-positives from T to C single nucleotide polymorphisms between the reference genome and HEK293 cell genome. Clusters with at least one nucleotide overlapping with repetitive elements were filtered out with the exception of low complexity repeats for HuR, which contain T-rich and AT-rich sub-categories that are descriptive of the putative HuR RRE (see Figure 1). Genomic coordinates of repetitive elements were obtained from http://www.repeatmasker.org.

Motif finding. All groups mapping to mRNA and not overlapping repetitive regions were ranked using CLI scores. Listed below are the full cERMIT output (highlighted in bold are motifs presented in main text results) and parameters applied. cERMIT is available at http://www.genome.duke.edu/labs/ohler/research/transcription/cERMIT.

cluster	motif	score(#targets)
1	**WTTTA**	**34.13(69881)**
2	**YTTTA**	**32.57(67708)**
3	**ATTTY**	**28.82(72755)**
4	WTTAA	19.32(41152)
5	TTAAW	19.16(38342)
6	TWTAA	19.05(41359)
7	YTTAA	18.77(40917)
8	TYTAA	18.55(40305)
9	AWTTA	16.74(34103)
10	WAATT	15.99(34663)
11	ATWTA	15.27(32433)
12	WTAAT	15.27(29505)
13	TAATW	15.22(29263)
14	AATWT	14.99(38239)
15	AATTY	14.56(34362)
16	ATTYA	14.49(31705)
17	AATYT	14.14(34426)
18	AYTTA	13.85(31842)
19	WTAAA	12.32(26287)
20	TAAAW	12.31(23736)
21	TATYA	11.62(23337)
22	WAAAT	11.35(22844)
23	YATTA	11.14(24096)
24	ATTAY	11.08(24719)
25	AYTAT	10.97(24076)
26	TYAAA	10.53(25906)
27	YTTCA	10.24(30499)
28	ACTTY	10.24(27429)
29	ATATY	10.13(26041)
30	AAAYT	9.45(20974)
31	AAATW	9.00(23940)
32	AATWA	8.61(17098)
33	WAAAA	6.05(14467)
34	AACWT	6.04(16889)

geneset scoring parameters
using regression scoring: no
strand-specific: no
p->degen_threshold: 0.25
p->fraction_degen_pos_threshold: 0.25
p->scoring_params.max_gene_set_size_percentage_threshold: 0.95
p->scoring_params.min_gene_set_size_absolute_threshold: 1000
p->scoring_params.min_gene_set_size_percentage_threshold: 0.05
p->min_motif_length: 5
p->max_motif_length: 10

ensemble PSSM parameters
p->required_core_length: 5
p->pssm_crop_threshold: 0.50
p->cluster_pssm_threshold_fraction: 0.50
p->fraction_of_top_sequences_to_consider: 0.50
species_type_global: human

clustering similar motif predictions parameters
p->cluster_sim_threshold: 0.99
p->hypegeom_p_val_cutoff: 1.000000e-30

MISC parameters
estimate p-value: yes

qRT-PCR

iScript (Bio-Rad) was used to make cDNA. Roche Lightcycler with Sybr green detection (Invitrogen) was used for qPCR and $\Delta\Delta$CT analysis method, using GAPDH for normalization. Primers for mature mRNA were intron spanning. All pre-mRNA specific primers were designed with forward primer in the intron preceding the 3' UTR and the reverse primer in the 3' UTR. No reverse transcriptase controls were run for PCR of pre-mRNA products. These either did not amplify at all or amplified no less than 10 cycles after the quantified product. All amplification products were of the expected size and melting temperature.

Mature mRNA primers:

GAPDH_fwd	CATTGCCCTCAACGACCACTTTGT
GAPDH_rev	TCTACATGGCAACTGTGAGGAGGG
TNF_fwd	AGGGACCTCTCTCTAATCAGCCCT
TNF_Rev	GTTATCTCTCAGCTCCACGCCATT
CCNA2_fwd	TGCTGGAGCTGCCTTTCATTTAGC
CCNA2_rev	TTGACTGTTGTGCATGCTGTGGTG
ACTB_fwd	ACCAACTGGGACGACATGGAGAAA
ACTB_rev	TAGCACAGCCTGGATAGCAACGTA

Pre-mRNA specific primers:

preBRCA1_fwd	GGTGTGATGGCATGTGCCTGTAAT
preBRCA1_rev	ATCTGCCCAATTGCTGGAGACAGA
preNFATC3_fwd	AGCCATGGGAAGGGAAATGTCTGA
preNFATC3_rev	TTGGAAACCCAAGGTCCAAGGAGA
preMYBL2_fwd	ACCAGGGTCTGTTGGGAACACATA
preMYBL2_rev	ACCCTCAACACCTCAGGACAAGAT
prePKM2_fwd	ACATTTCAGGGCCTTCCTCCTGTT
prePKM2_rev	TCAGCACAATGACCACATCTCCCT
preMDM2_fwd	TGCTAGCATTCCTGTGACTGAGCA
preMDM2_rev	GGCCCAACATCTGTTGCAATGTGA
preCTCF_fwd	TTGACTGTCTCTGGACCGCTATCT
preCTCF_rev	CTGTTGCTGGCAAAGAAGAGCACA

Comparison of spatial patterns

Distance between binding sites

Each cluster's nearest companion cluster was calculated and compared to a null in which the same number of clusters are randomly chosen and distributed in their genic region (e.g. intron, 3’ UTR…) with a minimum distance of 11nt between each cluster. Observed density estimates and density estimates of the 1st, 50th, and 99th percentile of the null were displayed.

Distance from transcript landmarks

The distance of each cluster from the 5' and 3' of each genetic region was compared to a null in which the clusters are randomly distributed in their genic region. For intronic spatial analysis the median density of the null permutations was then subtracted from the observed density to emphasize enrichment. In 3' UTRs that contain at least one Ago and HuR cluster, the position of the closest HuR cluster to each Ago cluster was calculated and summed across all 3' UTRs.

Ago and HuR clusters were randomly repositioned within their 3' UTRs to generate null distributions. Observed density estimates and density estimates of the 1st, 10th, 50th, 90th, and 99th percentile of the null were displayed.

Positional patterns and exon usage

For each exon, the number of upstream and/or downstream HuR binding sites were determined and merged with statistically significant HuR knockdown derived exon expression scores determined by FIRMA based on mapping common ENSEMBL exon IDs. For individual binding sites, upstream/downstream annotation was determined by distance to the nearest exon. Exons were divided into four categories: upstream HuR binding sites, downstream HuR binding sites, both upstream and downstream HuR binding sites, and no adjacent HuR binding sites. We then compared the density estimates of the FIRMA scores for each category. HuR RNA splicing maps were created (as described in experimental procedures) for exons annotated constitutive or not-constitutive by AltAnalyze. Similar analysis was performed for differentially expressed vs not-differentially expressed exons.

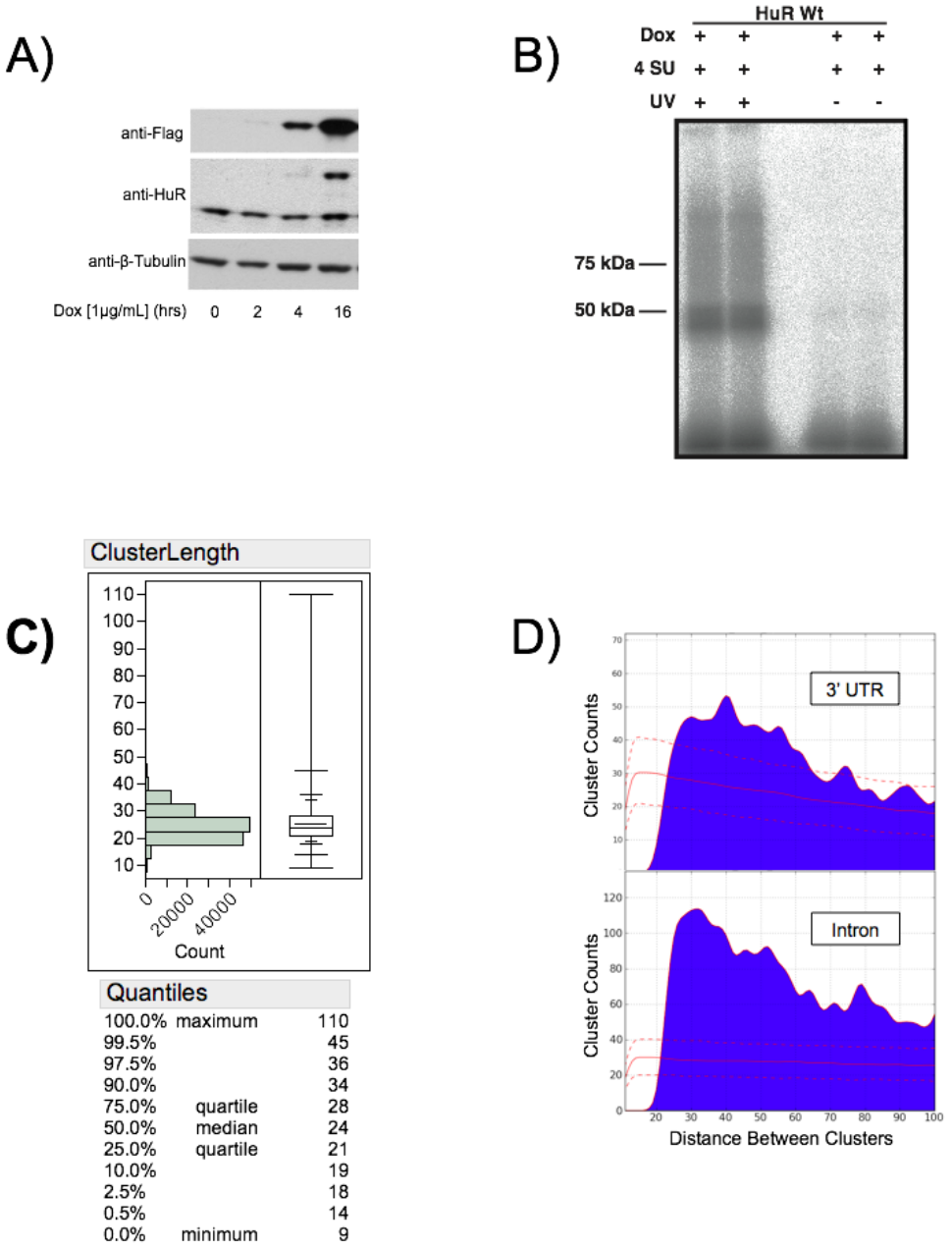


Figure S1. HuR PAR-CLIP Characteristics, Related to Figure 1. A) Doxycycline induction timecourse for FLAG-HA-HuR show approximately equivalent amounts of epitope tagged HuR and endogenous HuR after overnight induction. B) A radioactive band representing directly bound RNA fragments at expected size of the epitope tagged HuR (~45 kD). C) Histogram and quantile plot of HuR binding site lenghts shows most clusters are smaller than 30-nucleotide from the CCR method (Hafner et al). D) HuR binding sites in both introns and 3' UTRs are more likely to be near each other than random (red dashed lines represent 1st, 50th, and 99th percentiles of the null distribution).

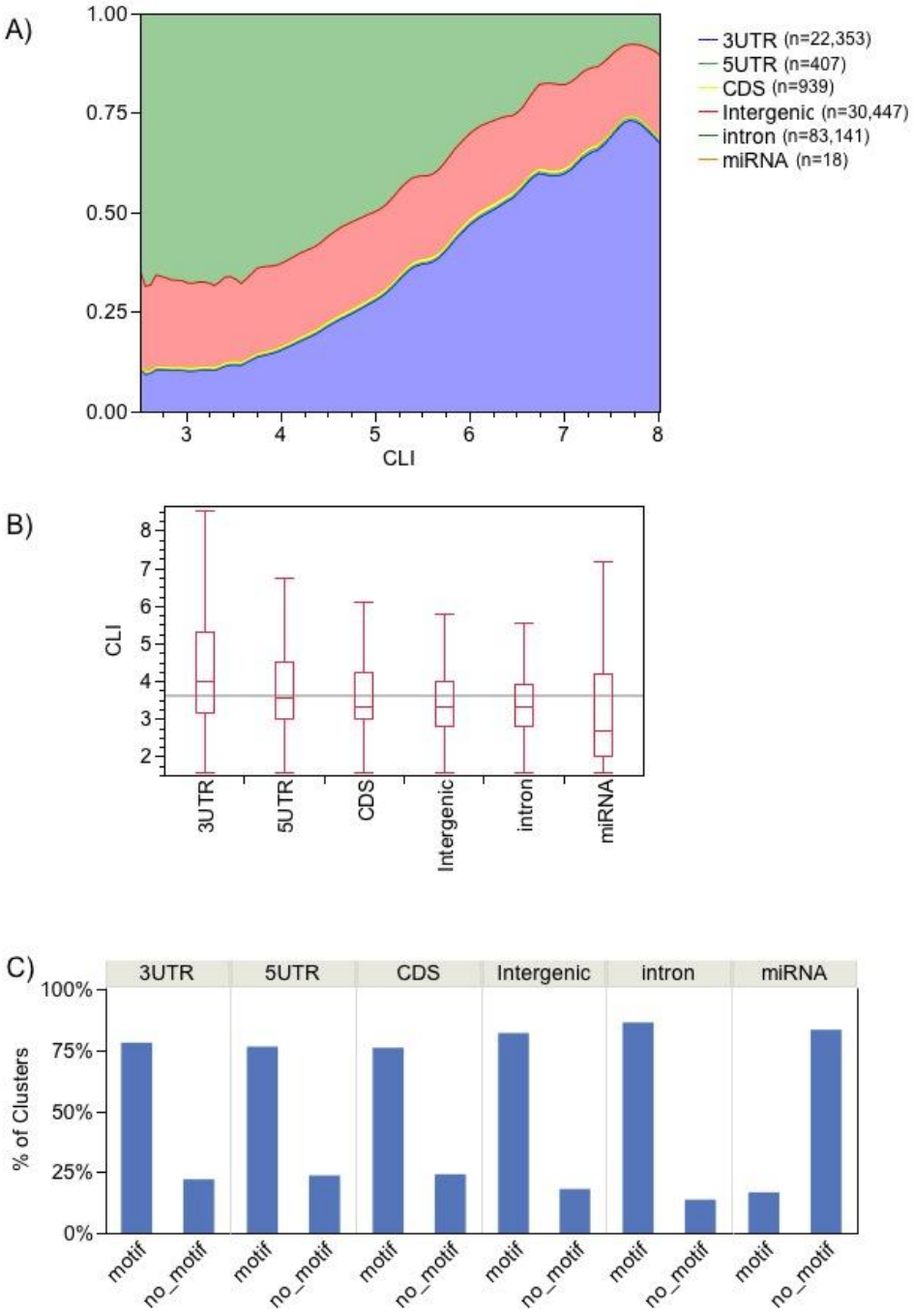
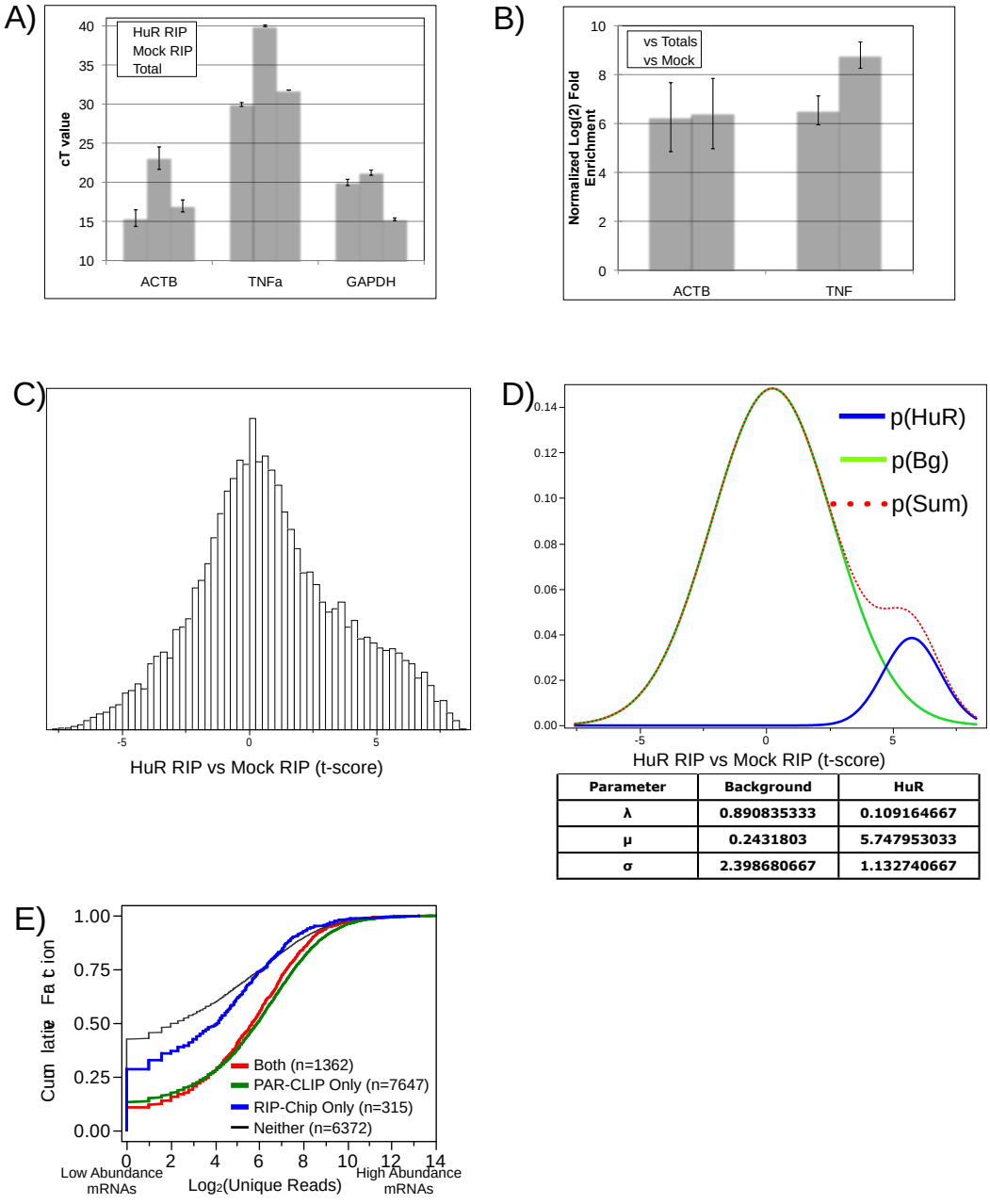
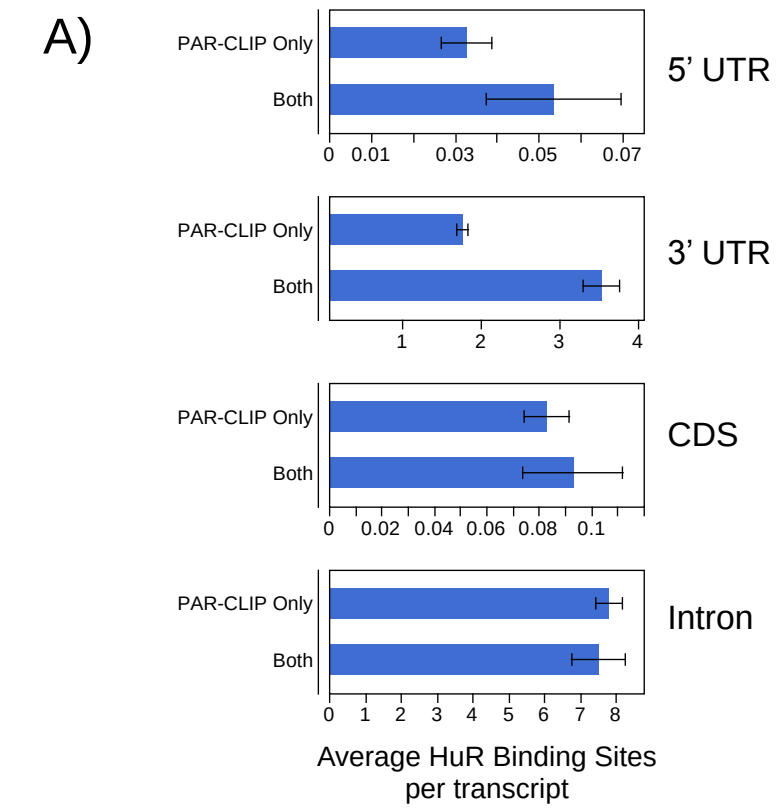


Figure S2. Landscapes of HuR Binding Sites in the Transcriptome, Related to Figure 2. Clusters were categorized based on whether they derived from intron, 5' UTR, CDS, 3'UTR, intergenic or miRNA regions. A) The proportion of densities and B) a quantile plot of CLI scores are displayed separately for each category. There is a clear hierarchy in the CLI scores for clusters derived from each region. C) Clusters from all regions were equally likely to contain on of the three identified HuR motifs, except for those derived from miRNAs for which there were very few clusters to begin with.



Parameter	Background	HuR
λ	0.890835333	0.109164667
μ	0.2431803	5.747953033
σ	2.398680667	1.132740667

Figure S3. HuR RIP-Chip, Related to Figure 3. A) qRT-PCR of signal of controls from biological triplicate RIPs. **TNF was not detected in mock IP and therefore the cT was set to 40, which is the total number of cycles. This value was used to calculate enrichment as well. B) Significant enrichment of positive controls over both mock and totals. C) Distribution of t-score comparison of triplicate HuR RIP vs Mock RIP. D) Gaussian mixture model of t-scores using 2 components (model parameters listed in box below). E) Comparison of the cumulative distribution of transcript abundance for targets detected by RIP-chip only (blue), PAR-CLIP only (green), and RIP-chip and PAR-CLIP (red) using RNA-seq data (Hafner et al) to quantify transcript abundance.



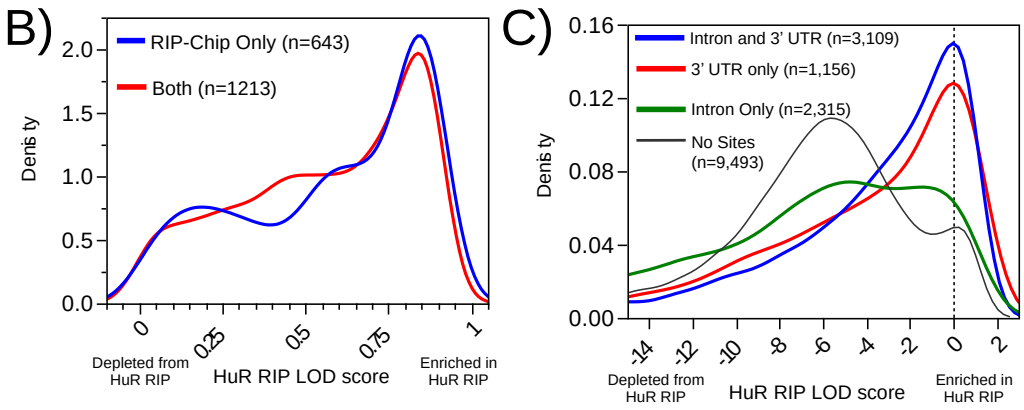
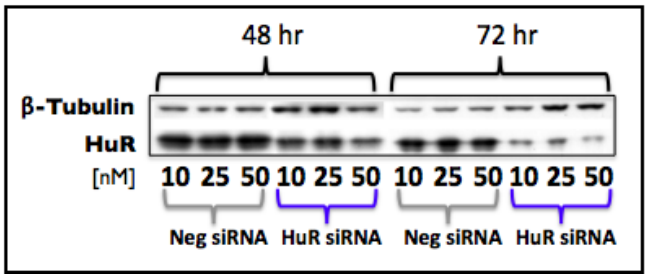
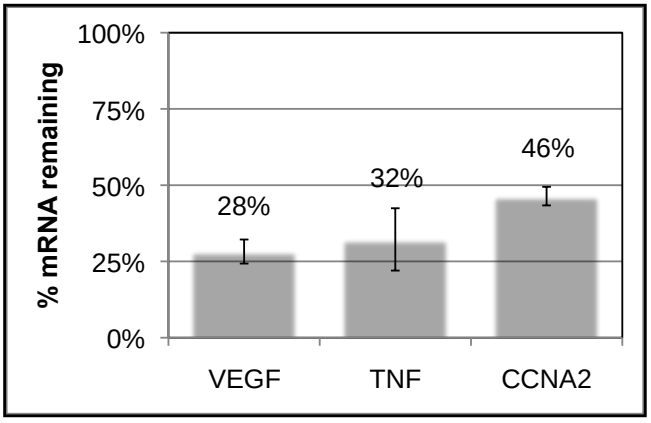


Figure S4. Comparison of PAR-CLIP and RIP-Chip Data, Related to Figure 4. A) Comparison of the number of binding sites in distinct regions of the mRNA for targets identified by both PAR-CLIP and RIP-chip or PAR-CLIP only (error bars represent 95% CI). B) Comparison of the distribution of HuR LOD scores for mRNA targets identified by RIP-chip only, or both RIP-CHIP and PAR-CLIP. C) Comparison of the distribution of HuR LOD scores for mRNAs containing binding sites only in 3' UTRs, only in introns, and in both 3' UTRs and introns, as well as those without binding sites.

A)



B)



C)

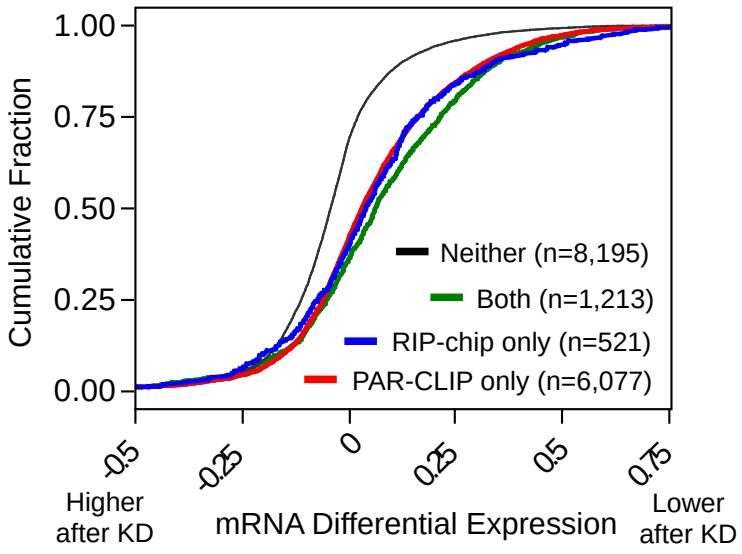


Figure S5. siRNA Depletion of HuR, Related to Figure 5. A) Three concentrations and timepoints of HuR siRNA knockdown were tested previous to transcriptomic analysis. B) Abundance of known HuR targets decrease after siRNA-mediated HuR depletion. The error bars represent standard error. C) Cumulative distribution of mRNA differential expression after HuR knockdown for targets identified by RIP-chip only (blue), PAR-CLIP only (red), and both (green), and neither (black).

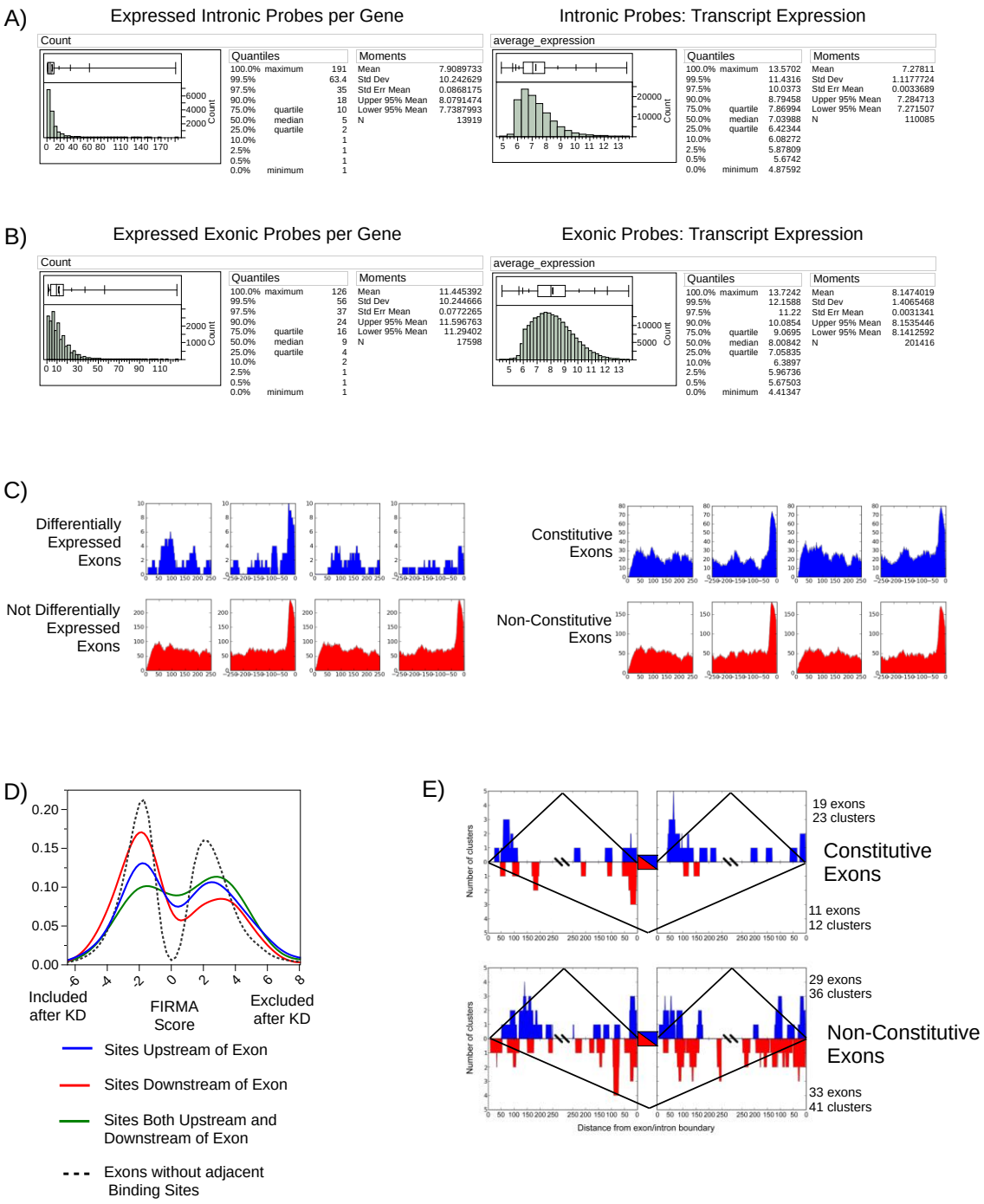


Figure S6. HuR and Pre-mRNA Processing, Related to Figure 6. Distributions of the number of probes/gene and median collapsed $\log_2$ gene expression for probes mapping to either A) intronic or B) exonic (5' UTR, CDS, 3' UTR) regions. C) Overrpresentation of HuR binding sites within 50 nts upstream of the 5' splice site. D) Relationship between position of intronic HuR binding sites and exon usage. Exons were classified into four categories: upstream HuR binding sites, downstream HuR binding sites, both upstream and downstream HuR binding

sites, and no adjacent HuR binding sites. The distribution of FIRMA scores was depicted for each exon category. E) HuR RNA splicing maps for exons labeled constitutive or not-constitutive. Combined with Figure 6E, these data suggest HuR promotes inclusion more often than exclusion for alternative (not-constitutive) exons more often than constitutive exons.

Table S1. PAR-CLIP Library Mapping Statistics, Related to the Experimental Procedures

(A)

HuR reads	19,594,213
HuR reads aligned to genome	4,718,747
HuR clusters	151,519
HuR clusters after filtering	137,305
HuR clusters mapping to a gene & pass filter	106,858

(B)

AGO = All 4 libraries combined	
AGO reads	19,572,167
AGO reads aligned	5,172,019
AGO clusters	37,328
AGO clusters after filtering	21,275
AGO clusters mapping to a gene & pass filter	18,492

For A) HuR library and B) Ago1-4 libraries combined.

Table S5. HuR and miRNA Binding Site Comparisons, Related to Figure 5

HuR depletion	
Comparison	p-Val
Overlap vs No_Overlap	1
Overlap vs +miR, -HuR	0.1133
Overlap vs -miR, +HuR	0.9951
Overlap vs -miR, -HuR	<.0001
No_Overlap vs +miR, -HuR	0.0001
No_Overlap vs -miR, +HuR	0.7983
No_Overlap vs -miR, -HuR	<.0001
+miR, -HuR vs -miR, -HuR	0.0003
-miR, +HuR vs +miR, -HuR	0.0003
-miR, +HuR vs -miR, -HuR	<.0001

miRNA depletion	
Comparison	p-Val
Overlap vs No_Overlap	**0.0005**
Overlap vs +miR, -HuR	0.0015
No_Overlap vs +miR, -HuR	0.999
-miR, +HuR vs Overlap	**0.8714**
-miR, +HuR vs No_Overlap	<.0001
-miR, +HuR vs +miR, -HuR	<.0001
-miR, -HuR vs Overlap	**0.1076**
-miR, -HuR vs No_Overlap	<.0001
-miR, -HuR vs +miR, -HuR	<.0001
-miR, -HuR vs -miR, +HuR	<.0001

Student's t-tests were used to calculate p-values for each pairwise comparisons made in Figure 7C and D.